

UNITED STATES
DEPARTMENT OF THE INTERIOR
GEOLOGICAL SURVEY

TECHNICAL LETTER NUMBER 2

DIGITAL PROCESSING
OF ARRAY SEISMIC RECORDINGS

By

Alan Ryall* and John Birtill**

August 10, 1962

DENVER, COLORADO

This page intentionally left blank

UNITED STATES
DEPARTMENT OF THE INTERIOR
Geological Survey

Technical Letter
Number 2

Dr. Charles C. Bates
Chief, VELA UNIFORM Project
Advanced Research Projects Agency
Defense Department
Pentagon
Washington 25, D. C.

Dear Dr. Bates:

Transmitted herewith are 10 copies of:

TECHNICAL LETTER NUMBER 2

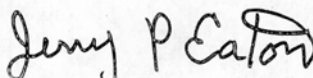
DIGITAL PROCESSING
OF ARRAY SEISMIC RECORDINGS

By

Alan Ryall* and John Birtill**

August 10, 1962

Sincerely yours,



J. P. Eaton
Acting Chief
Branch of Crustal Studies

* U. S. Geological Survey, Denver, Colorado

** United Kingdom Atomic Energy Authority, Reading, England.
(Mr. Birtill was in England at the time this technical letter was prepared, and was not available to review the entire manuscript)

ABSTRACT

This technical letter contains a brief review of the operations which are involved in digital processing of array seismic recordings by the methods of velocity filtering, summation, cross-multiplication and integration, and by combinations of these operations (the "UK Method" and multiple correlation).

Examples are presented of analyses by the several techniques on array recordings which were obtained by the U. S. Geological Survey during chemical and nuclear explosions in the western United States.

Seismograms are synthesized using actual noise and P_n -signal recordings, such that the signal-to-noise ratio, onset time and velocity of the signal are predetermined for the synthetic records. These records are then analyzed by summation, cross-multiplication, multiple correlation and the UK technique, and the results are compared.

For all of the examples presented, analysis by the non-linear techniques of multiple correlation and cross-multiplication of the traces on an array recording are preferred to analyses by the linear operations involved in summation and the UK Method.

I. INTRODUCTION

In seismological studies which involve the use of single-seismometer recordings, arrivals or phases can be detected only by changes in the character of the recording (i.e., a change in amplitude or frequency or both). The detection or identification of seismic signals which are obscured by noise can be improved by arranging seismometers in groups or arrays. On array recordings, in addition to changes in signal character on the individual traces, the criterion of coherence of the signal from trace to trace can be used for identification. To develop the usefulness of the array to its full potential, techniques of analysis which take into account various known, or anticipated, properties of the signal or noise (cross-correlation, velocity filtering, noise prediction, etc.) can be applied to the records.

The theoretical problem of enhancing a signal by combining in various ways the output of individual detectors in an array has been treated in connection with specific communication processes (Kraus, 1950) and also in general communication theory (Lee, 1960; Blackman and Tukey, 1958; Berman and Clay, 1957). Detailed theoretical considerations of array processing as related to the specific problem of seismic-signal detection were recently presented as a VELA UNIFORM report to the Air Force Technical Applications Center by Texas Instruments Incorporated (1961). The various theoretical studies indicate several approaches which might be fruitful in identifying seismic events on array recordings, but it is difficult to predict, on purely theoretical grounds, the improvement which would be obtained in applying a given technique to a given group of data.

The usefulness of one type of processing, or the advantage of one method over another, is dependent upon several factors: the nature of the signal, the nature of the noise, and the coherence of the noise from one detector to another. On records of blasts made by the U. S. Geological Survey, these parameters are practically unpredictable and vary considerably from one recording location to another.

Applications of array-processing techniques to seismic recordings of earthquakes and blasts have been limited. A method that involves velocity filtering, summation and cross-correlation of array signals has been used recently by workers of the United Kingdom Atomic Energy Authority, and this method has produced results which seem to represent a striking improvement in the detection of events on array seismograms.

This technical letter is a preliminary report on the digital-processing techniques that have been applied to date on records obtained by the U. S. Geological Survey. The present results are still too sketchy and few in number to lead to the selection of a single method or combination of methods in the routine processing of our records. Some of the techniques have significantly improved our ability to pick phases on noisy records and to identify weak or mixed signals, however, and our work to date has indicated the direction in which we should proceed in developing a practical system of digital processing of array seismograms.

II. DIGITAL PROCESSING OF ARRAY RECORDINGS

We shall consider a linear array of N evenly-spaced seismometers, making an angle α with the direction toward the source. If the spacing between the seismometers is d , a wave traveling outward from the source with apparent velocity v and arriving at the first seismometer at time t_0 will reach the i -th detector at time $t = t_0 + (i-1) d \cos \alpha / v$. The output of the array will consist of N time functions, $f_i(t)$, $1 \leq i \leq N$, each containing the desired signal plus unwanted noise.

If the frequency content of the signal is approximately known, then some initial improvement of the signal-to-noise ratio of the recording can be obtained by filtering the $f_i(t)$, using a pass-band equal to the anticipated frequency range of the signal. We shall designate the filtered traces as $g_i(t)$.

After the initial filtering, the recording can be further improved in many instances by application of one or more of a number of linear and non-linear processes. One of these is the process of velocity filtering and summation of all the $g_i(t)$, which can be expressed by the equation

$$s_v(t) = \sum_{i=1}^N g_i(t + \delta t_i), \quad (1)$$

where $\delta t_i = (i - 1) d \cos \alpha / v$.

If the time functions $g_i(t)$ have mean values of zero and contain no long-period trends, then a second operation which can be used for signal enhancement is that of multiple correlation. In multiple correlation the traces are shifted and multiplied together, and the resulting time series is smoothed by integrating.

The multiplications are represented by the equation

$$p_v(t) = \prod_{i=1}^N g_i(t + \delta t_i), \quad (N \text{ even}) \quad (2)$$

and the smoothing is expressed by

$$u_v(t) = \sum_{j=0}^m p_v(t + j\Delta t), \quad (3)$$

where Δt is the interval used in digitizing the record and m is the length of integration.

In multiple correlation, if the $g_i(t)$ are coherent from trace to trace (and if N is even) the time series $p_v(t)$ will be positive, but for random noise or noise which does not correlate for the velocity v , the values of $p_v(t)$ will be both positive and negative. In the smoothing represented by equation (3), the part of $p_v(t)$ which is generated by noise will tend to cancel, provided m is taken large enough, but for a coherent signal of sufficient duration (i.e., a few cycles) $u_v(t)$ will attain large positive values.

The results of multiple correlation, in contrast to those of shifting and summation of the traces, can be rendered useless unless care is taken in the preparation and selection of the traces for analysis. A false "coherence" can be indicated if there is a lack of symmetry (or presence of a long-period trend) of one or more of the traces about its mean, or zero, value. In forming the products indicated by equation (2), a trace or part of a trace which contains extremely low amplitudes ("dead" trace) will cause extremely low

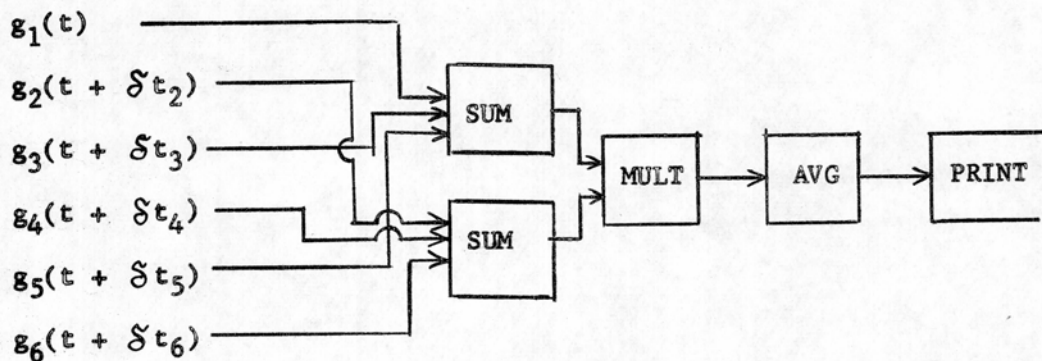
amplitudes to appear in $p_v(t)$, and as a result, in the final correlogram $u_v(t)$. A trace which is out of phase with the other traces (as with a very noisy trace) might obscure the coherence of all or part of the desired signal. When records are carefully selected and prepared for multiple correlation calculations, however, this technique appears to be very powerful in the identification of phases which are either obscured by high background noise or which contain amplitudes too small to be reliably identified on the original recording.

A third technique (the "UK Method") combines the two operations described above. The N traces are first separated into two groups, and the traces of each group are shifted and summed according to equation (1). The method of separation is arbitrary: group I, for example, might consist of traces 1 to $N/2$, or of all the odd-numbered traces. Since the seismic noise on our records is usually not coherent over distances of several hundred meters, the method of combining the odd- and even-numbered traces into two groups is preferred to a combination involving adjacent traces.

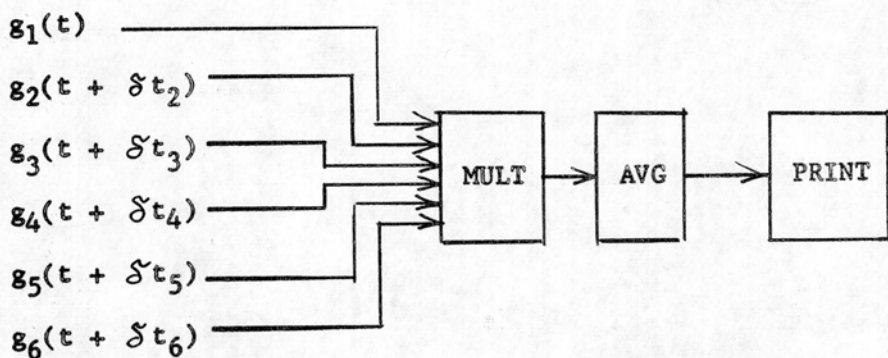
The shifting and summation operations result in two time series, $s_{1v}(t)$ and $s_{2v}(t)$, for a given velocity v . These two series are then cross-correlated, according to the equation

$$z_v(t) = \sum_{j=1}^M s_{1v}(t + j \Delta t) \cdot s_{2v}(t + j \Delta t). \quad (4)$$

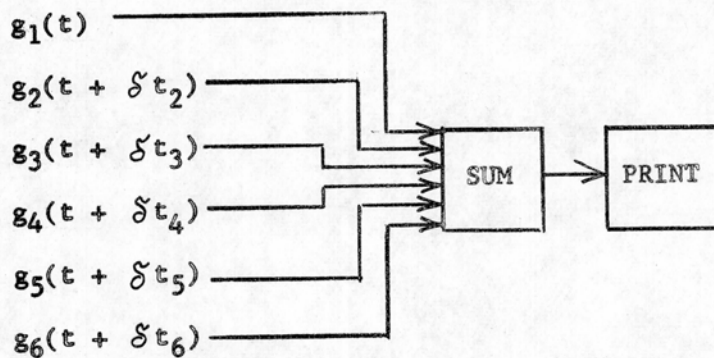
The UK Method has the advantage that the signal-to-noise problems mentioned above in connection with dead or noisy traces can be taken into account by summing before the cross-correlation. In a detection scheme, therefore, this method might be better suited for phase detection by routine analysis of array recordings, where the records are not first subjected to special de-trending, selecting, and filtering operations. Calculations of the multiple correlation type, however, have higher velocity discrimination than the UK Method (velocity discrimination is analogous to "directivity" in Berman and Clay, 1957), and, with some preliminary treatment of the data, might be preferred to the UK technique for routine phase detection. The sense of first motion of a phase (compression or dilatation) might be determined by simple summation of the shifted traces after the approximate arrival time of the phase had been identified by one of the other methods. Examples of the three types of operations are shown in figure 1, for a six-element array.



(a)



(b)



(c)

Figure 1. Examples of the operations described in section II.

(a) -- the UK Method; (b) -- multiple correlation;

(c) -- summation.

III. SELECTED RESULTS OF ANALYSIS

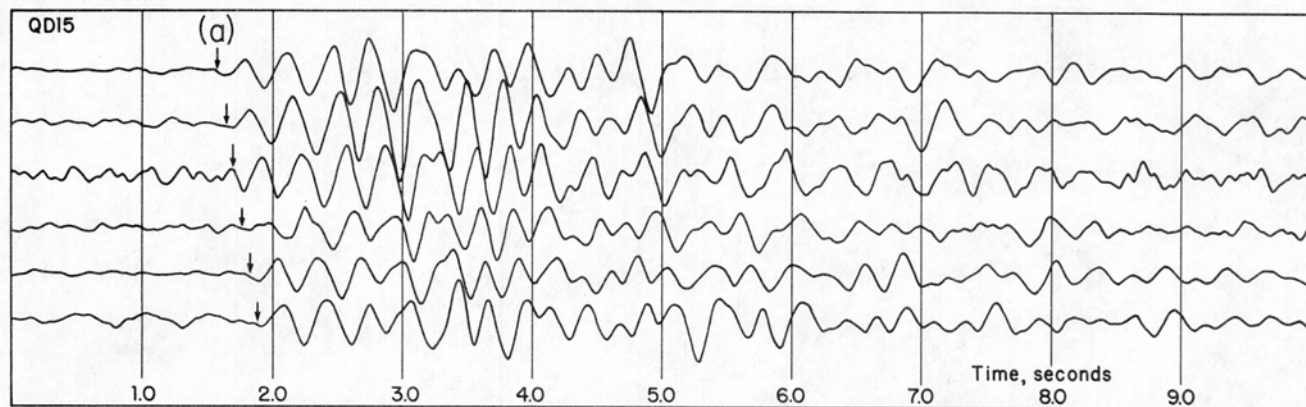
Results of analyses of several of the U. S. Geological Survey array recordings are shown in figures 2-9.

Much of the preliminary analysis and program development was carried out using digitized amplitudes of a record (QD15) which was obtained during the "HARDHAT" explosion (15 Feb., 1962) at a distance of 991 kilometers (near Pueblo, Colorado). Digitization of this record (and of the records described below) was accomplished using a photographic playback of the magnetic tape recording of the event; a sampling rate of 50/second was used. The portion of the seismogram which was digitized is shown in figure 2, together with the results of analysis by summation, multiple correlation and the U. K. technique. The point in time at which the P_n arrival was picked by eye is designated by "(a)" in the figure.

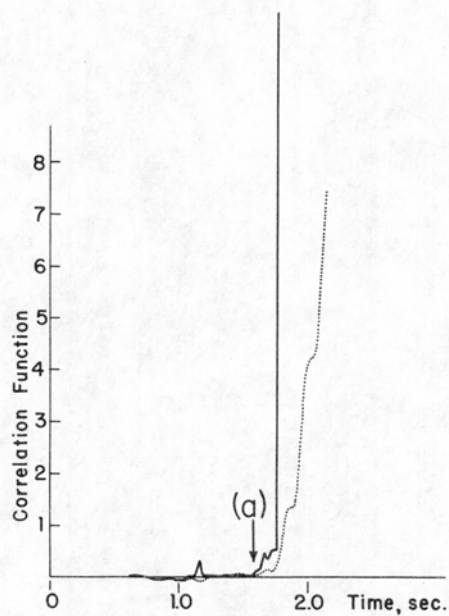
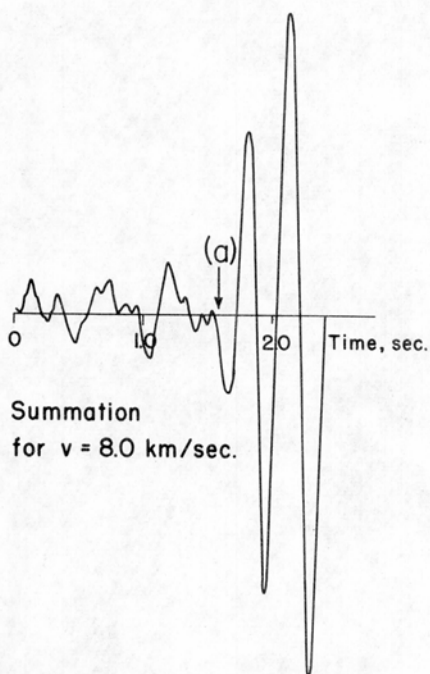
Several points should be mentioned regarding figure 2. The most important of these is that a greater portion of the background noise preceding the onset of the signal should have been included in the digital data. The lack of analysis of a sufficient quantity of background noise weakens the conclusions which follow, regarding the identification of a relatively long-period signal which precedes point (a) on the seismogram.

By shifting and summing the six traces of QD15 for an apparent velocity of 8.0 km/sec (lower-left plot in figure 2), the arrival at point (a) is sharply enhanced, and is seen to signal the onset of a 3-cps train which continues for about a second. There is an additional suggestion in this plot of a weak forerunner, which precedes point (a) and contains a frequency of about 1.8 cps. Such a forerunner, preceding the higher-frequency arrival by about a second, might be expected at this distance

DIGITAL ANALYSES OF AN ARRAY RECORD



Original recording ("HARDHAT" explosion, $\Delta = 991$ km.)



Analyses for $v = 8.0$ km/sec, by:

- the "U.K." Method
- Multiple Correlation

Figure 2.

on the basis of travel-time data at other recording locations.

However, the identification of such an arrival on the summation trace in figure 2 is questionable, due to the absence of a longer sample of background noise.

In the lower-right-hand corner of figure 2 are shown the results of calculations corresponding to equations (3) and (4) of section II. The "correlation function" shown on the ordinate of the plot represents relative values for $u_{8.0}(t)$ and $z_{8.0}(t)$ in equations (3) and (4), respectively. Both of the techniques have "picked" the 3-cps arrival at point (a), and the plot of $z_{8.0}(t)$ is seen to rise very sharply at a point corresponding to the crest of the second leg of this arrival. The sharp onsets and extremely large amplitudes obtained by multiple correlation will be shown below to be a function of the cross-multiplication of the six traces. These sharp onsets in correlation do not correspond in time to zero-crossings of the original record, but to times of maximum amplitudes of the signal.

The correlograms in figure 2 do not indicate the longer-period forerunner mentioned above in relation to the results of summation. The correlogram for $v = 9.5$ km/sec (not shown) does, however, indicate an arrival at $t = 1.05$ seconds. The amplitude of the correlation function, $z_{9.5}(t)$, for this arrival would have a value of 0.35 on the scale in figure 2, and this amplitude remains constant until the correlation function increases at point (a). For the reason stated above, this early pick is not considered to be overly reliable; a longer section of QD15 is presently being digitized in order to study the significance of the early arrival.

IDENTIFICATION OF WEAK ARRIVALS BY MULTIPLE CORRELATION

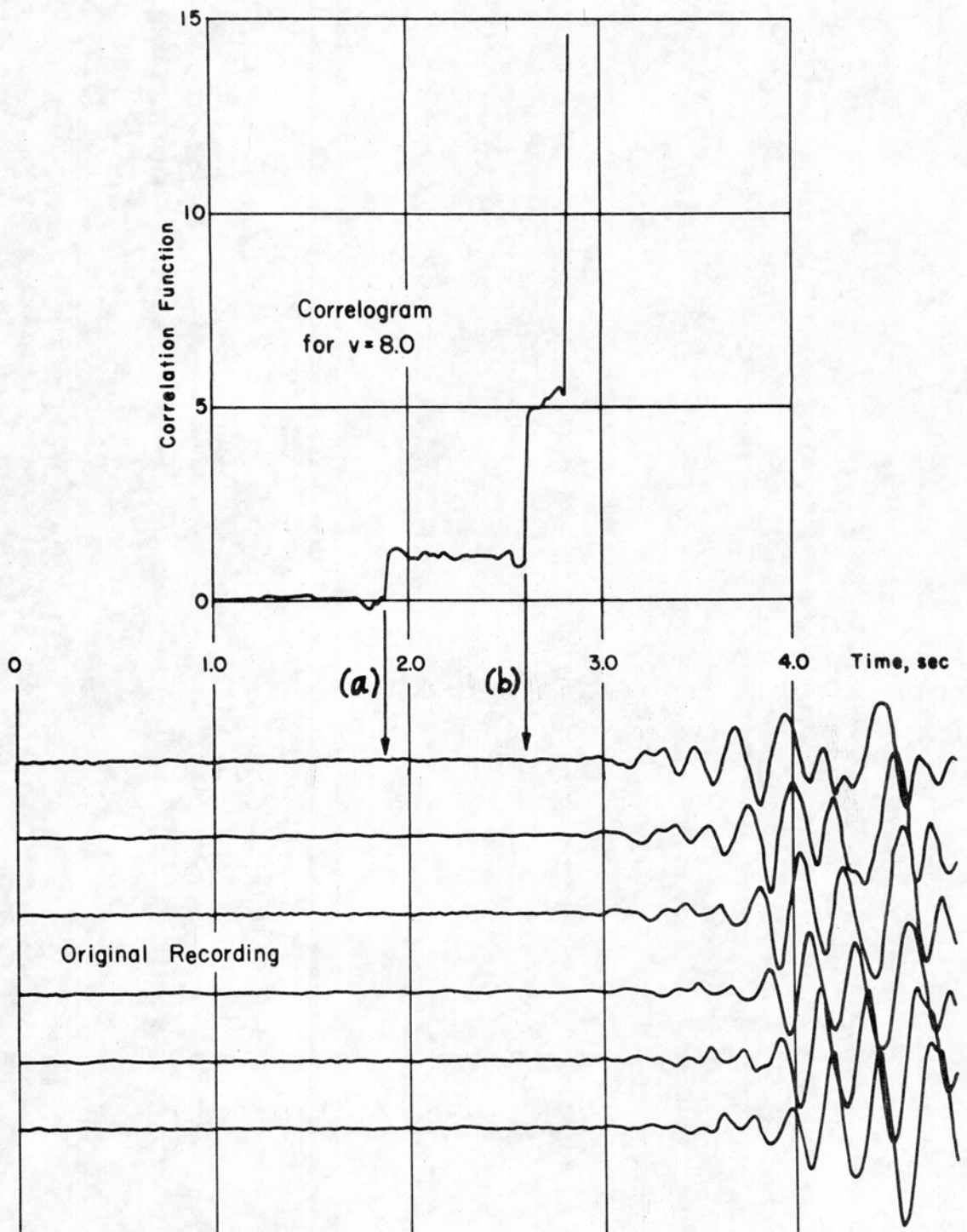
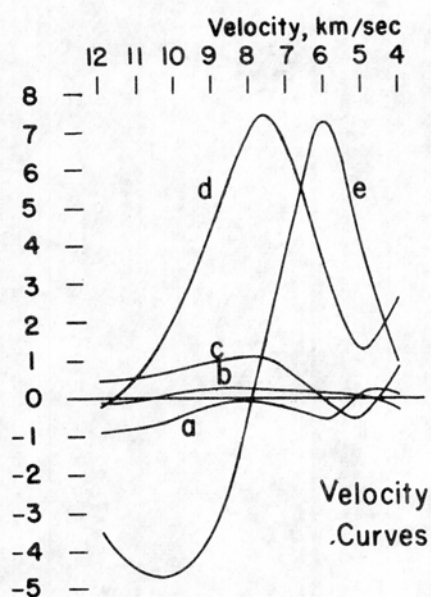


Figure 3.



IDENTIFICATION OF PHASES ON A NOISY RECORD BY MULTIPLE CORRELATION

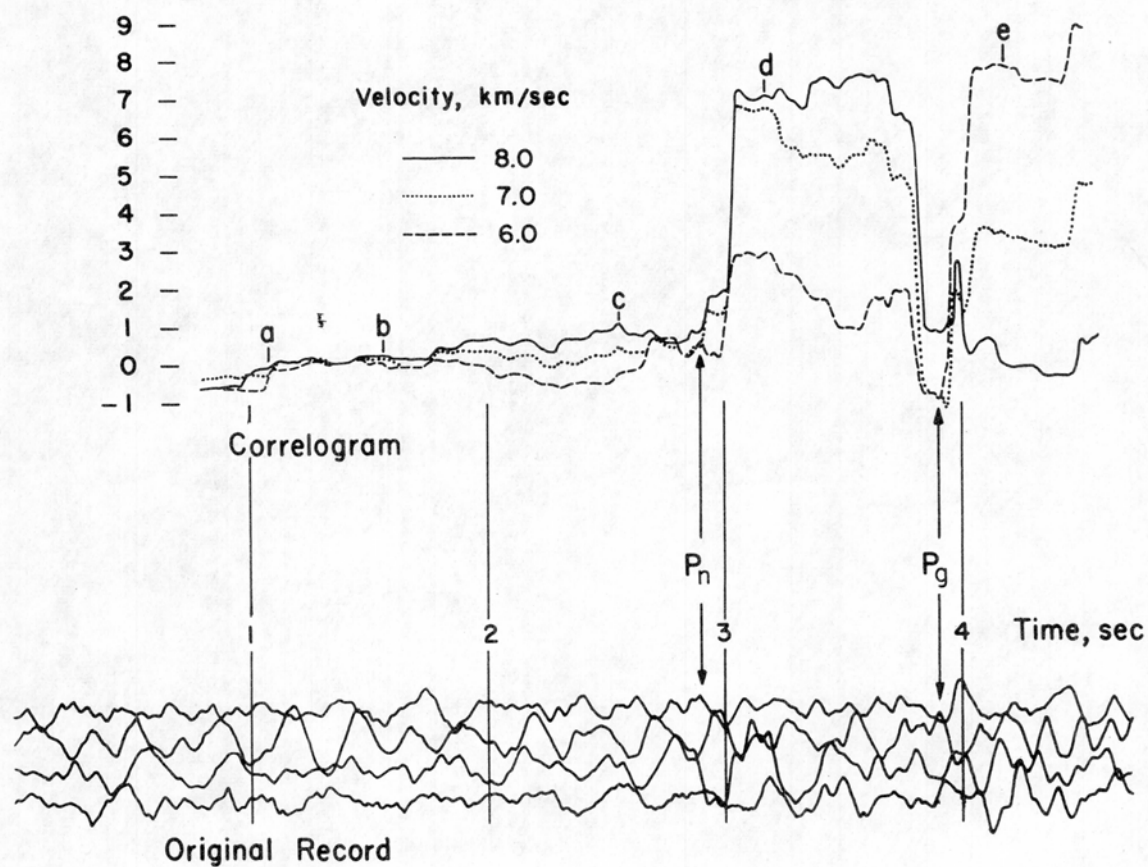


Figure 4.

The result of an attempt to identify weak arrivals by multiple correlation is shown in figure 3. The seismogram in this figure (JD15) was made during the "HARDHAT" explosion, at a distance of 618 kilometers (near Monticello, Utah). As with the previous example, an insufficient amount of background noise was included in the digital data, and the first arrival picked by the correlation program (point (a) on the correlogram) should be considered questionable. The sharp onset at (b) would correspond, on the travel-time curve, to the early pick -- the long-period forerunner -- on QD15. This arrival was identified by eye on the original recording before the multiple correlation calculations were made; it was, however, listed as a "very poor" pick. On the basis of the arrival time and distance for this explosion, for a short-period Benioff station at Durango, Colorado (supplied to the USGS by Dr. Carl Romney of AFTAC), the onset of P_n might be expected at $t = 2.2$ seconds on the record in figure 3 (Note: The times on this and other correlograms are referred to the first, or upper, trace of the seismogram. Since the value of m was taken to be 50 in the smoothing operation in equation (3), no values of the correlation function were calculated corresponding to the first second of time -- the first 50 samples -- of the original recording).

The results of an attempt to identify phases on a noisy record are shown in figure 4. The seismogram (SY6) in the lower portion of the figure was obtained near Salinas, California, during a chemical explosion near San Francisco. The distance from the shotpoint to the

first receiver (upper trace) was 112 kilometers. The times shown in the figure, for the original record and the correlograms, represent times along the digitized portion of the record; $t = 0$ in the figure would correspond to a travel-time of 17.37 seconds.

The digitized amplitudes of four traces of SY6 were analyzed by multiple correlation, for 20 velocities in the range $4.0 \leq v \leq 12.0$ km/sec. The resulting correlograms for three of these velocities (8.0, 7.0 and 6.0 km/sec) are shown in the center of the figure. It can be seen from this plot that two phases have been identified by the multiple correlation program: the first, which begins at about $t = 2.9$ seconds, is associated with the phase P_n , and the second, at $t = 3.9$ seconds, is presumed to be a phase of the type P_g . The correlation in P_n is highest for the 8.0 km/sec correlogram, while for P_g the graph for $v = 6.0$ km/sec predominates.

In the upper portion of figure 4 the correlation function amplitudes have been plotted as a function of velocity, for five points in time designated by (a) - (e) on the correlogram. Each velocity curve was constructed by plotting the value of the correlation functions computed for the given time, for the 20 different velocities. From the velocity curves (d) and (e), the phases P_n and P_g are seen to peak sharply at velocities of 7.7 km/sec and 6.0 km/sec, respectively. The curves plotted for points (a), (b) and (c), within the background noise preceding P_n , are erratic and contain low amplitudes. The large negative lobe on curve (e) appears to have been caused by an accidental correlation of the noise which follows the onset of the P_g signal.

This example serves to illustrate an important point which was mentioned in section II: The method of multiple correlation cannot be relied upon to pick the first motion, or onset time, of a signal which is obscured by high background noise. The multiple correlation "pick" for P_n for SY6 is found to be about 0.4 second late with respect to the travel-time curve for the Salinas region which was constructed using data obtained at a large number of recording locations during several explosions near San Francisco. The onset time indicated by the correlogram for P_g cannot be evaluated, since the crustal structure in this region is extremely complicated and P_g cannot be traced smoothly from one recording location to another.

The final example to be discussed is presented in figures 5-9. In the previous examples, an attempt was made to identify signals in the presence of noise on actual recordings, where little was known beforehand regarding the nature of the desired signals or their actual arrival times. For the present case, a large sample of recorded noise was mixed with a P_n signal, in a manner such that the onset time and velocity of the signal, as well as the signal-to-noise (S/N) ratios of the composite recordings, were predetermined.

Figure 5 illustrates the original noise recording, the signal, and portions of the composite records for three selected values of S/N. The composite records were constructed as follows:

(1) The six traces of noise (obtained during recording in Nevada) were digitized using a sampling interval of 0.02 second. Each of the noise traces was reduced by the mean value of its digitized amplitudes, and the noise-level of that trace was taken to be the root-mean-square (rms) value of the reduced amplitudes. These operations are given by the equations

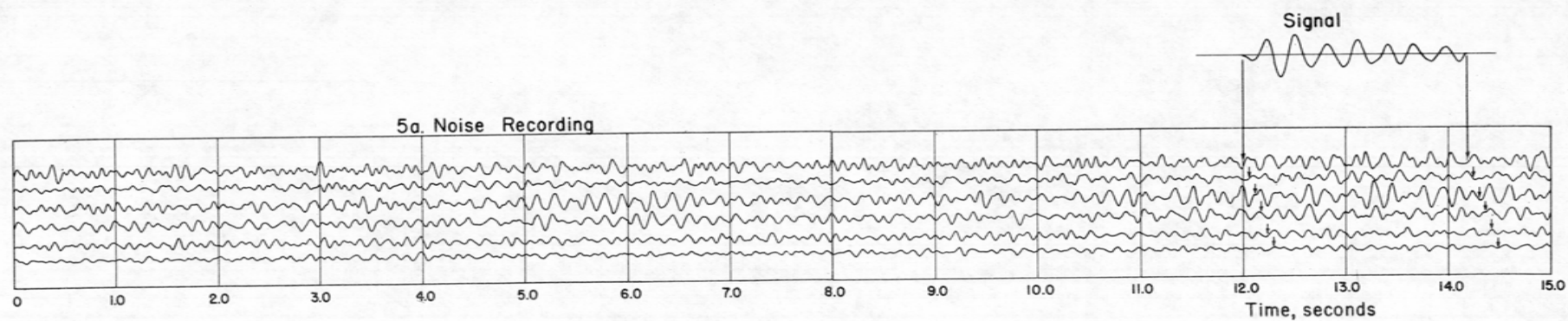
$$A_{ij} = a_{ij} - \frac{1}{m} \sum_{i=1}^m a_{ij}, \quad 1 \leq j \leq 6, \quad (5)$$

and

$$N_j = \left(\frac{1}{m} \sum_{i=1}^m A_{ij}^2 \right)^{1/2}, \quad (6)$$

where a_{ij} are the initial digitized amplitudes, A_{ij} are the reduced amplitudes, m is the total number of samples on a single trace ($m = 750$), and N_j is the noise-level of the j -th trace. The noise level of the 6-trace recording was then taken to be the average value of the N_j :

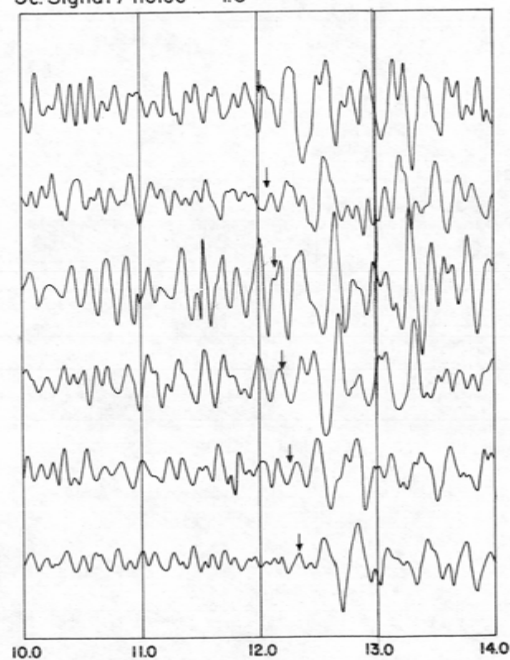
$$N = \frac{1}{6} \sum_{j=1}^6 N_j. \quad (7)$$



5b. Signal / noise = 2.0



5c. Signal / noise = 1.0



5d. Signal / noise = 0.5

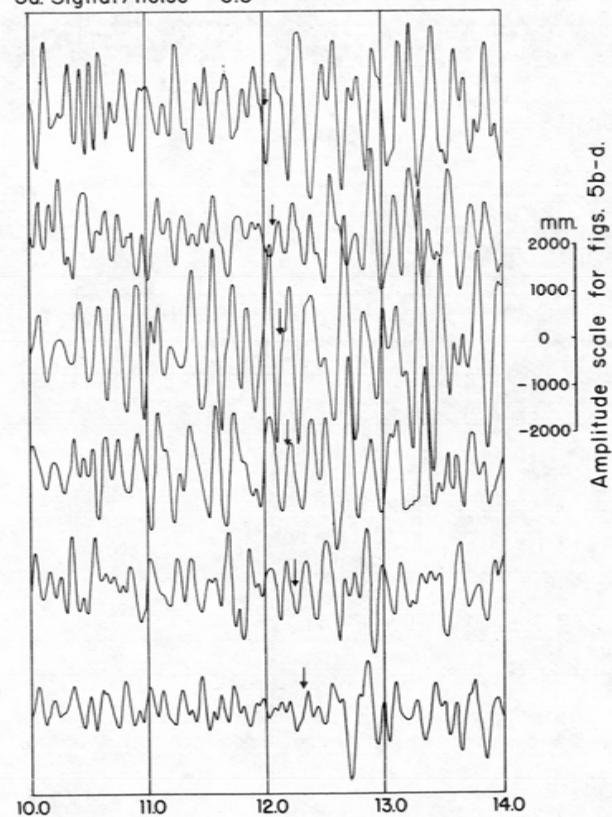


Figure 5.

(2) A P_n signal, obtained near Saguache, Colorado ($\Delta = 930$ kilometers), during the "HARDHAT" explosion, was digitized over a 2.2-second interval. The signal was reduced by its mean value, and the signal-level S was taken to be the rms value of the reduced signal amplitudes, S_i :

$$S = \left(\frac{1}{n} \sum_{i=1}^n S_i^2 \right)^{1/2} \quad (8)$$

(3) The ratio S/N was calculated and factors c_k were computed such that

$$\frac{S}{c_k N} = C_k \quad (9)$$

for values of C_k equal to 2.0, 1.0 and 0.5. For a given value of C_k , the amplitudes of the six noise traces were multiplied by c_k , and the signal (which was kept at a constant level) was added to the noise record so that its arrival time on the first trace was equal to $t = 12.02$ seconds, and its apparent velocity across the array was $v = 8.3$ km/sec.

In figure 5a the original noise and signal samples are shown, and in 5b-d portions of the composite records are presented for the three ratios of S/N . In all of the figures, the point of onset of the P_n signal on each trace is indicated by an arrow.

Several characteristics of the records in figure 5 should be noted:

(1) The signal selected for this example is a rather typical P_n signal, and begins in an emergent fashion. The rms value of the first half-cycle is less than the rms value S of the entire signal, by a factor of 2.5, while the rms value of the third half-cycle is greater than S by a factor of 1.6, and the peak amplitude of the third half-cycle is about equal to $2.4S$. Because of these relationships, as will be seen below, the onset of P_n is obscured by noise in the optimum case, when $S/N = 2.0$, while the third half-cycle of signal can be identified by multiple correlation for all three cases.

(2) The noise-levels of the individual traces vary considerably. The third trace is the noisiest and the sixth is the quietest, with $N_3 = 2.6 N_6$.

(3) The signal and noise were mixed on a portion of the record where the noise level is higher-than-average on traces 1-4. This tends to make the values of S/N somewhat smaller over that portion of the record containing signal-plus-noise.

(4) On none of the composite records, figures 5b-d, can the onset of P_n be identified with confidence. In figures 5b and 5c, some of the signal character has been preserved, while in 5d the signal is, for all practical purposes, completely obscured by noise on traces 1-5, and could be identified on trace 6 only with the previous knowledge that the trace contained a signal.

Figures 6-9 illustrate the improvement which can be obtained for the composite records of figure 5, by summation, cross-multiplication, the U. K. technique and multiple correlation.

In figure 6, the traces of each of the composite records have been shifted for a velocity of 8.3 km/sec, and the amplitudes summed. The values of S/N to the right of the summed traces are those of the original records; the amplitude scale is consistent with the scale shown for figures 5b-d. (Note: the same holds for figures 7, 8 and 9; however, the amplitude units in these figures are wrong, and should read mm^6 , mm^2 and mm^6 , respectively).

On none of the summed traces can the onset of P_n (arrows) be identified with certainty. All of these traces, however, contain frequency and amplitude characteristics of the original signal which are easily differentiated from those of the noise.

In figure 7, the six traces of the record for $S/N = 0.5$ have been shifted and multiplied together according to equation (2), for $v = 8.3$ km/sec. In plotting the amplitudes of $p_{8.3}(t)$, the ordinate values have been connected to the abscissa by straight lines, rather than connecting the ordinate values to one another by a smooth curve. This method of plotting tends, perhaps, to overemphasize the spiky nature of the results of cross-multiplication; however, a smooth curve through the points would bear a very close resemblance to the spikes in the figure. The spiky appearance of this plot can be easily seen to be due to the process of cross-multiplying the six traces, if one considers the result of taking the sixth power of several cycles of a sinusoid.

Improvement of a noisy record by summation

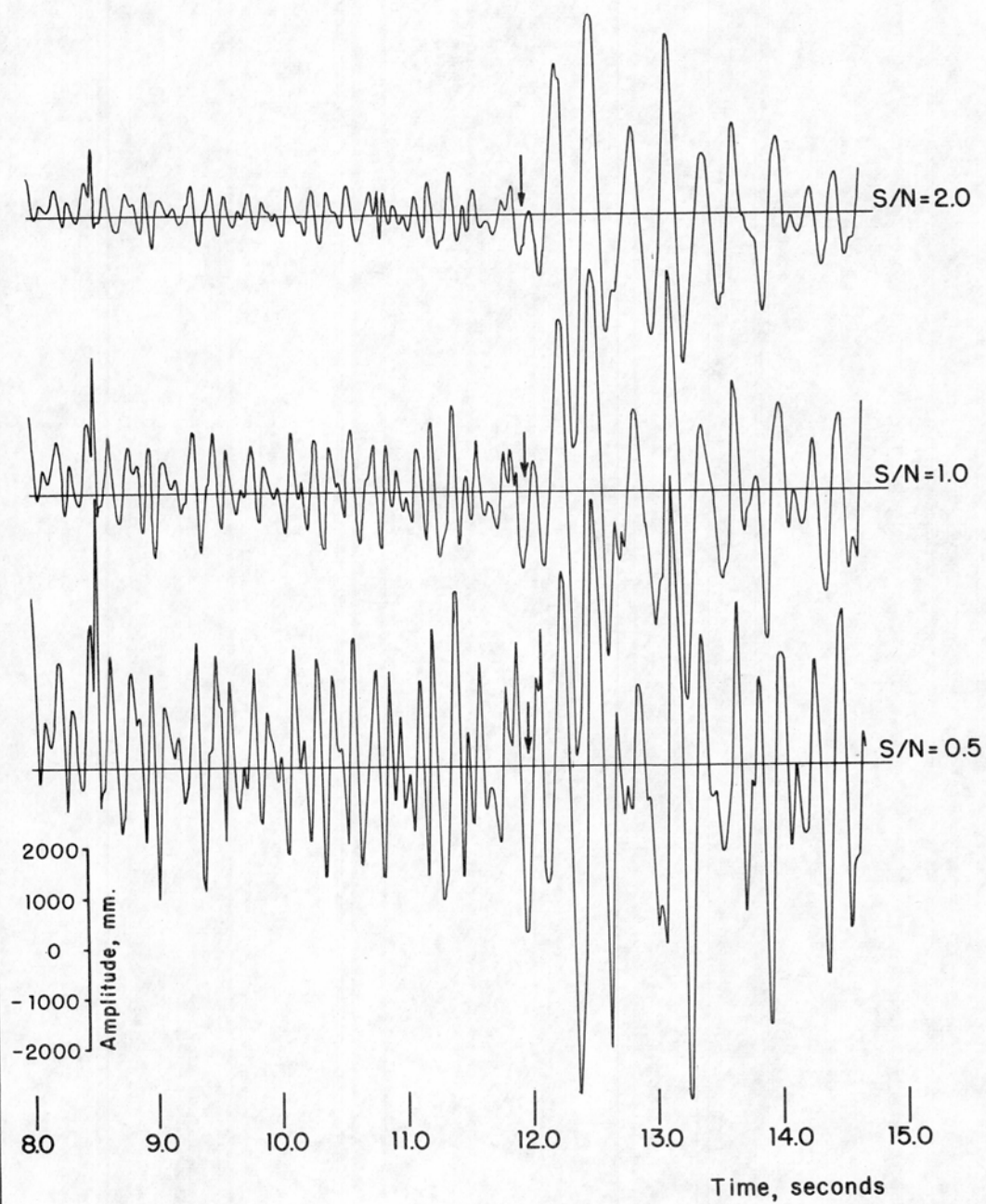


Figure 6.

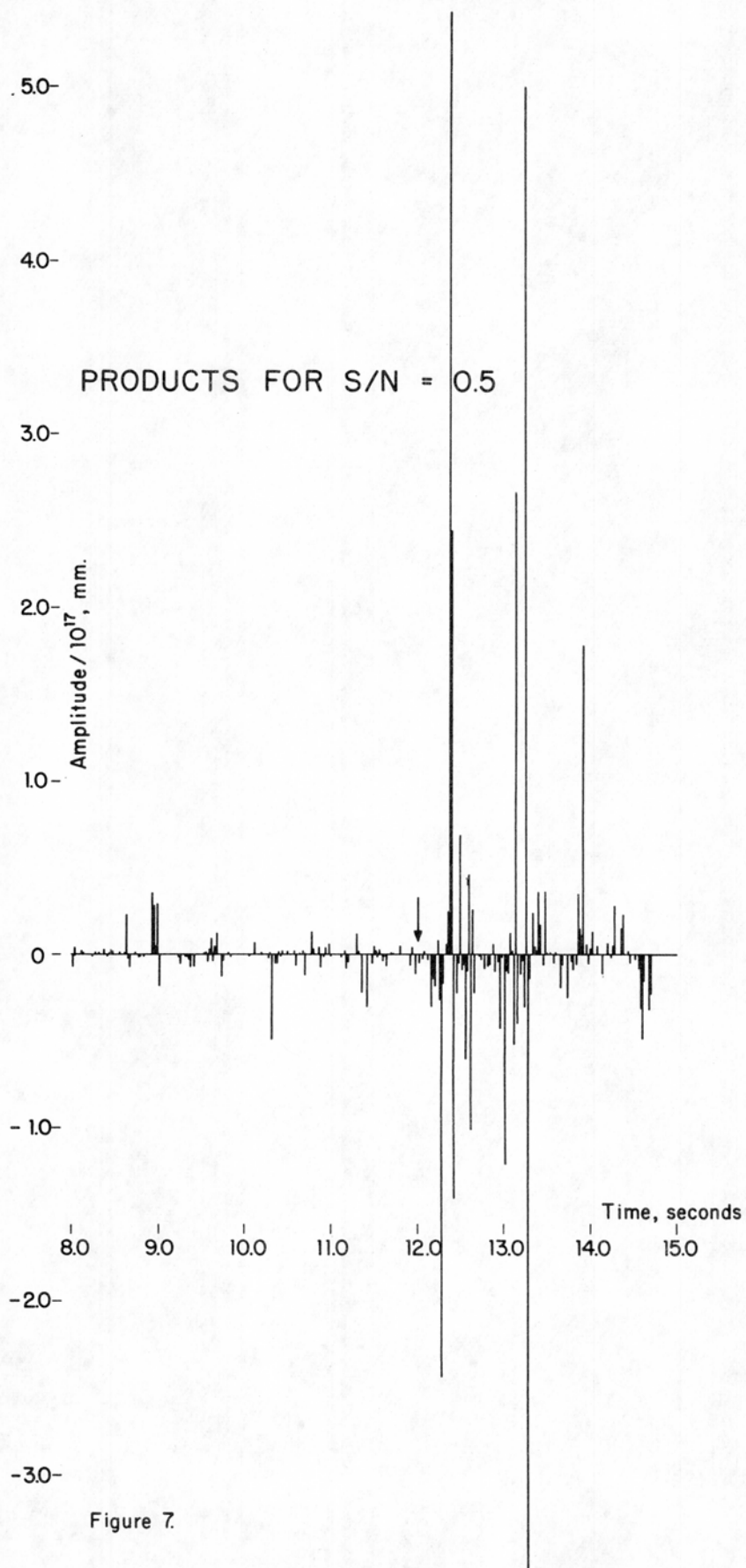


Figure 7.

The signal in figure 7 can easily be distinguished from the noise by a comparison of amplitudes, and the signal portion of this trace can be more easily identified (either by eye, or using a computer to compare amplitudes) than that of the lower trace in figure 6. Two disadvantages of the method of cross-multiplication should be noted: (1) frequency characteristics of the original signal are not preserved in this process, and (2) the high level of noise has caused much of the signal portion of the trace to have negative values. This is in contrast to the statement in section II that if the $g_i(t)$ are coherent from trace to trace, and the number of traces is even, the time series $p_v(t)$ will be positive. For automatic, on-line processing of continuous array recording, however, the technique of cross-multiplication might have advantages, both in cost and in presentation of the final trace, over the other techniques discussed in this paper.

The results of analysis by the UK technique and by multiple correlation are shown in figures 8 and 9. From these figures it can be seen that both of the analyses have clearly identified the maximum amplitude of the third half-cycle of the signal, at $t = 12.38$ seconds, for all three ratios of S/N . In addition, the maximum amplitude of the second half-cycle, at $t = 12.26$ seconds, was identified by both methods for the record $S/N = 2.0$, and by the UK technique for $S/N = 1.0$.

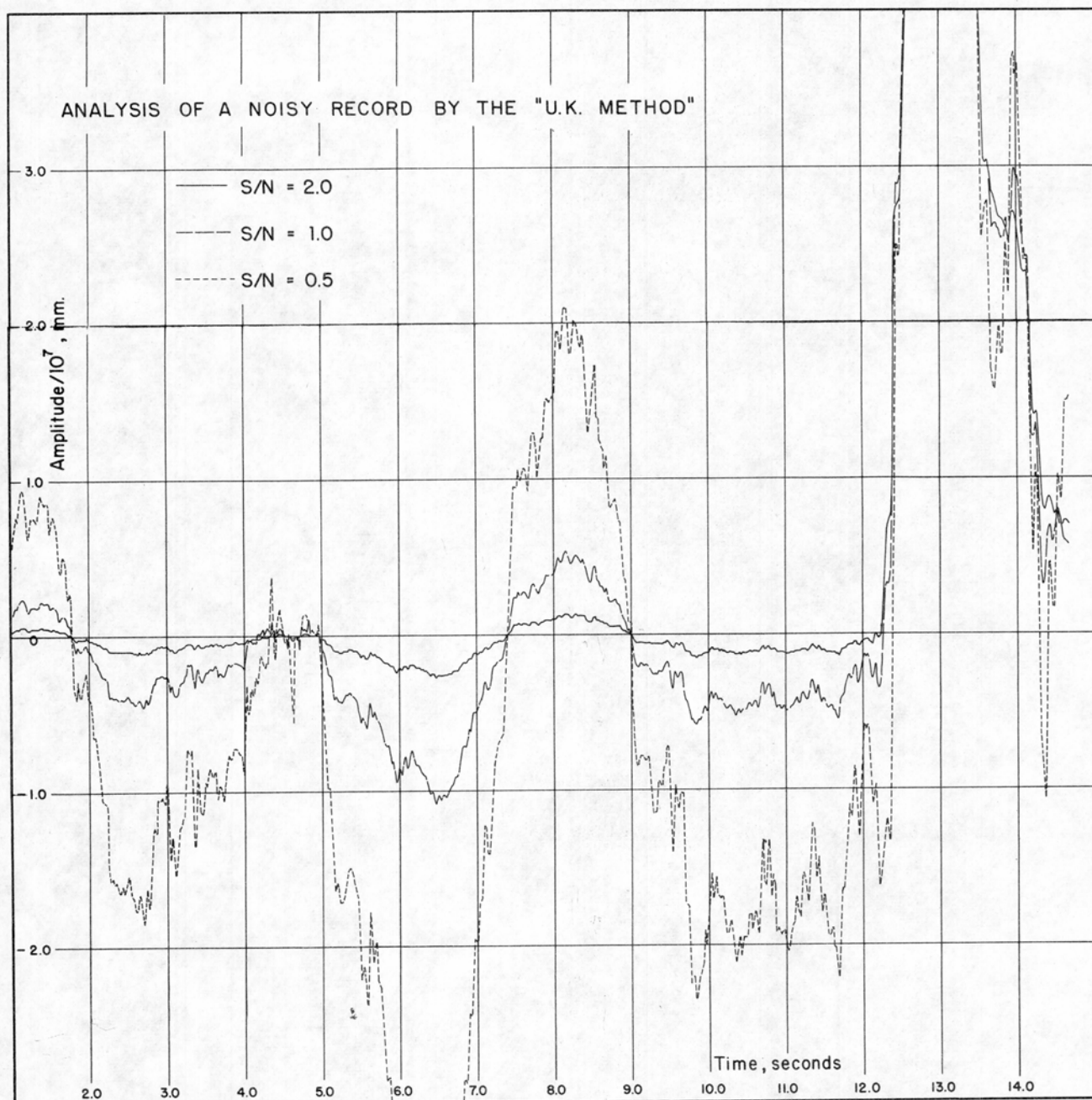


Figure 8.

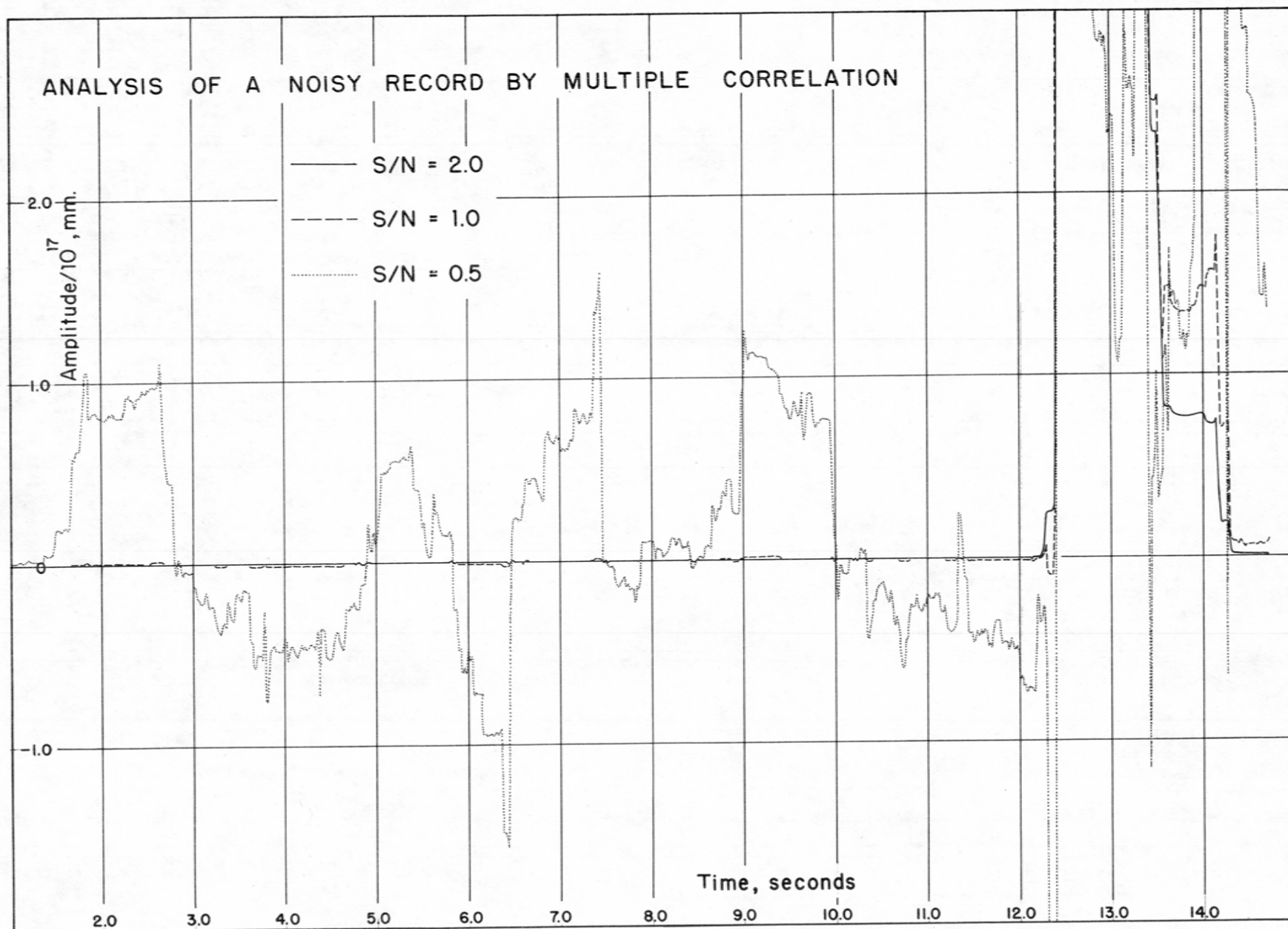


Figure 9.

For a comparison of the signal-to-noise improvement of the two methods, the maximum amplitudes of the signal and noise and the ratio of these amplitudes are presented in the following table (the maximum signal amplitude in all cases occurred at $t = 13.32$ seconds; on figure 8 the maximum noise amplitude occurs at $t = 6.62$ seconds; on figure 9 the largest value of the noise is at $t = 6.44$ seconds):

S/N of orig. record	Maximum signal amplitude	Maximum noise amplitude	Ratio of max. signal to max. noise
I. Analysis by the UK Technique			
2.0	6.2×10^7	-2.6×10^6	23.8
1.0	7.1×10^7	-1.0×10^7	7.1
0.5	8.7×10^7	-4.3×10^7	2.0
II. Analysis by Multiple Correlation			
2.0	7.2×10^{17}	-3.7×10^{13}	1.9×10^4
1.0	9.5×10^{17}	-2.4×10^{15}	4.0×10^2
0.5	6.5×10^{17}	-1.6×10^{17}	4.1

From a comparison of the tabulated amplitudes, as well as a comparison of the general appearance of the plotted curves in figures 8 and 9, it is concluded that, for the three records under consideration, analysis by multiple correlation is preferable to that by the UK method, for identification of the P_n signal. The latter technique was, however, capable of identifying an earlier portion of the signal for the record $S/N = 1.0$, and

from a consideration of the ratios in the fourth column of the table, this technique might be preferred to multiple correlation for phase identification where the signal-to-noise ratio of the original recording is less than 0.5. It should also be noted that the maximum signal amplitude on figure 9, for the record $S/N = 0.5$, would have been considerably higher if the large negative signal amplitudes obtained in cross-multiplication (figure 7) had been rectified before integration.

REFERENCES

1. Berman, A. and C. S. Clay "Theory of Time-Averaged-Product Arrays" Jour. Accoustical Soc. Amer., Vol. 29, No. 7, pp. 805-812, July, 1957.
2. Blackman, R. B. and J. W. Tukey The Measurement of Power Spectra. Dover Publications, Inc., New York, 1958.
3. Kraus, John D. Antennas. McGraw-Hill, New York, 1950.
4. Lee, Y. W. Statistical Theory of Communication. John Wiley and Sons, New York, 1960.
5. Spieker, L. J., Project Manager. "Seismometer Array and Data Processing System" Texas Instruments, Inc., Project VELA UNIFORM, AFTAC Project VT/077, Final Report, 1961.